

# HMRC VAT update December 2015

Author: Mark Stephens and John Allcoat, Office for National Statistics

Date: 21 December 2015

## Executive summary

This is the latest in a series of updates on the work to utilise data collected by Her Majesty's Revenue and Customs (HMRC) from Value Added Tax (VAT) returns as an administrative data source for Short-term Output Indicators (STOI) and National Accounts. The STOI in scope are the [Index of Production \(IoP\)](#), [Index of Services \(IoS\)](#) and [Output in the Construction Industry](#).

This article:

- describes the plan to use HMRC turnover data as part of an estimate of nominal gross domestic product (GDP) and the strategic implications this has for National Accounts (section 2);
- describes the Monthly Business Survey (MBS) and considers how this would be impacted by the move to a new nominal estimate of GDP in the medium-term (section 3);
- shares raw experimental HMRC VAT turnover data at industry level, comparing with MBS data at industry level (section 4);
- describes our aim to pilot changes commencing summer 2016 to Index of Services and the Output approach to measuring GDP by ceasing MBS for elements of some industries and replacing these with data from HMRC turnover (section 4.2);
- provides an update on further progress with the HMRC VAT project (section 5).

This is the third in a series of articles; the first article was published on 14 August 2015 entitled [Feasibility study into the use of HMRC turnover data within Short-term Output Indicators and National Accounts](#) with the second published on 7 October 2015 entitled [Exploitation of HMRC VAT data](#).

## 1 Background

The [first article](#) set out the broad plans and intentions for the project and explored some of the benefits and challenges. While the [second article](#) provided more detail on our processing of data from HMRC and described some additional benefits surrounding the use of HMRC data, the "[Spending Review 2015](#)" and the "[Independent Review of UK Economic Statistics: Interim Report](#)" set a clear path to improve the use of administrative data against a backdrop of improved efficiency whilst improving the quality of our data. In this context the latter states that

"monthly VAT returns to HMRC on turnover and purchases of individual businesses, matched through an improved business register, has the potential to greatly enhance the output measure of GDP" (p20).

## 2 The strategic sourcing of HMRC data

### 2.1 HMRC turnover

In the previous two articles we described how turnover data from HMRC offers the potential for qualitative improvements to National Accounts. This is primarily due to the significantly improved coverage of business turnover activity which HMRC data represents. At present the ONS Monthly Business Survey (MBS) collects turnover data from a census of large businesses, and a randomly selected sample of small and medium-sized businesses in selected industries. In addition, other industries such as oil and gas extraction and air transport are measured through the use of volume data. At this point we should be clear that appropriate volume data will continue to be collected in future as it is available in a more timely fashion than data collected from HMRC.

However, in future HMRC turnover data offers the potential to achieve a census of turnover activity for smaller and medium-sized businesses in industries covered by the MBS. While it may also deliver turnover estimates for industries that are currently measured using volume sources. We would then be able to sum business turnover, similar to output, using appropriate census data from MBS and HMRC at real or nominal values – that is including inflation.

### 2.2 HMRC expenditure

We have also previously described how data collected by HMRC through a VAT return on business purchases could be used to model intermediate consumption - as the purchases data also includes expenditure on fixed assets (intermediate consumption is the value of the goods and services consumed as inputs by a process of production, excluding fixed assets).

Nominal values of output and intermediate consumption are at present derived primarily from the ONS [Annual Business Survey](#) and through the ONS supply and use balancing framework. The supply and use tables used in the framework bring together information from the expenditure, income and output approaches to measuring GDP and derive the value-added, that is, output less intermediate consumption, which connects each approach. We would then be able to derive value-added at nominal values.

This methodology for deriving value-added currently only applies when the data are available to populate the supply and use tables – largely through the less timely annual surveys. In the long-term, the strategic sourcing of data from HMRC may allow this method to be used in a more timely basis. In addition, we would also be able to take account of the different inflation rates that impact output and intermediate consumption by deflating each dataset with the appropriate price deflators to improve the volume estimate of value-added – a practice known as double deflation.

### 3 Understanding the Monthly Business Survey

#### 3.1 Background of the MBS

We have explained how data from HMRC may have a strategic significance in the long-term development of National Accounts and how in future estimates from MBS for small and medium-sized businesses could be replaced by data from HMRC. The MBS is and will remain a vital source of data and it would be helpful to understand its methodology in more detail.

The [MBS](#) commenced in 2010 with the merger of the Monthly Production Inquiry and the Monthly Inquiry into the Distribution and Services Sector which trace their modern origins to 1991 ([The development of the Monthly Business Survey](#)) The MBS for October 2015 was despatched to 32,300 businesses representing 43% of GDP in 2012 which is the latest year for which GDP weights are calculated. The coverage of MBS is best described in the [GDPO source catalogue](#) which also highlights how various sources such as government and household final consumption expenditure are used as current price data series in the compilation of short-term indicators. The survey is split into five sampling stratum based on the registered employment of each business on the [Inter Departmental Business Register \(IDBR\)](#). Band 1 represents the smallest businesses in a given industry and the sample is selected at random from the universe of businesses contained on the IDBR for the sampling period. Bands 2 and 3 similarly represent small and medium-sized businesses which are selected at random although for a few small industries Band 3 can also be a census of activity. While Band 4 represents the largest businesses in a given industry, and the selection constitutes a census of businesses in excess of the employment cut-off for that industry. In addition, businesses below the census cut-off with turnover in excess of £60m can represent a significant part of an industry and so these undergo a forced selection at Band 5. There are relatively few of these businesses.

It should be made clear that although Band 1 selections for many industries are confined to the 0-9 employment size-band, for some industries Band 1 covers 0-4 employment. Similarly Bands 2 and 3 can differ in make-up between different industries in terms of the employment size-bands used to delineate the Bands.

### 3.2 MBS selection for October 2015

In the Production industries, [UK 2007 Standard Industrial Classification \(SIC 2007\)](#) divisions 07-39, the MBS covered 6,000 businesses and the Band 4 cut-off is primarily at 50, 100, 150 or 250 employment. For the period October 2015 the selected businesses covered 74.8% of the registered turnover of the industries selected on the IDBR. Table 1 considers the sample in more detail at Band level. It is notable that the 74.8% coverage is achieved through 71.9% of that figure from Bands 4 and 5 and an additional 2.9% from Bands 1 to 3. It also highlights that 1,000 forms for Band 1 industries cover only 4.3% of the registered turnover universe.

**Table 1 – Characteristics of the MBS October 2015 sample for the Production industries**

	Percentage share of universe turnover	Forms despatched	Percentage share of sample
Band 4/5	71.9%	2,900	48.6%
Band 3	15.0%	1,100	18.4%
Band 2	8.8%	1,000	16.5%
Band 1	4.3%	1,000	16.5%
<b>Total</b>	<b>100.0%</b>	<b>6,000</b>	<b>100.0%</b>

In the Services industries, SIC 2007 divisions 45-96, the MBS covered 26,300 businesses and the Band 4 cut-off is primarily at 100 employment. For the period October 2015 the selected businesses covered 70.5% of the registered turnover of the industries selected on the IDBR with Bands 4 and 5 totalling 67.8% and Bands 1 to 3 an additional 2.7%. Table 2 considers the sample in more detail and it should be noted that Band 1 businesses constitute a larger part of the sample than in the Production industries – but this is due to the greater significance that Band 1 industries have in representing 10.3% of the registered turnover universe.

**Table 2 – Characteristics of the MBS October 2015 sample for the Services industries**

	Percentage share of universe turnover	Forms despatched	Percentage share of sample
Band 4/5	67.8%	10,900	41.3%
Band 3	12.1%	4,800	18.4%
Band 2	9.8%	4,300	16.2%
Band 1	10.3%	6,300	24.1%
<b>Total</b>	<b>100.0%</b>	<b>26,300</b>	<b>100.0%</b>

For completeness it should be added that the MBS is not restricted to identifying total turnover in a period. Every end-quarter month (March, June, September, December) an employment question is asked as part of the [Business Register Employment Survey \(BRES\)](#). In addition, in the manufacturing industries (SIC 2007 10-33), MBS collects data on export turnover.

### 3.3 Changes to MBS: a pilot approach

As previously stated, the MBS samples for Bands 1 to 3 are a random sample of businesses. For some industries the samples achieved by MBS are small – sometimes as few as 10 businesses per Band for the smaller industries. In comparison the HMRC data offers a far greater coverage. Over the coming years we will explore the use that can be made of HMRC turnover data for Bands 1, 2 and 3. In this context, it is notable that the smallest businesses in Band 1 are estimated with the despatch of 1,000 MBS forms for the production industries and 6,300 MBS forms for the services industries while in comparison at 6 months after the reference period some 898,000 businesses reported HMRC turnover within Band 1. (This figure has been calculated from the November 2015 vintage of HMRC turnover and is illustrative of the amount of reporting units for June 2015.)

We believe that the scale of this dataset has significant potential and will allow us to slim the number of forms despatched to businesses – particularly to Band 1 but also to Bands 2 and 3. However, it must be made clear that we do not believe a simplistic or blanket approach is warranted. Decisions on the use of HMRC turnover data and the potential to cease the despatch of MBS forms will be made on an industry by industry basis. The sample for Band 4 will not be impacted by the potential use of HMRC VAT turnover.

We believe that a pilot approach is required to test the working assumptions and performance of the new estimates. We have a significant time-series commencing in January 2007 at our disposal to test the data and importantly we have frozen the VAT datasets at every month from January 2014. This means we can parallel-run the data at each month and judge the impact of the new method with new data every month – nearly 2 years of parallel-runs. This testing will allow the flexibility to refine our method before the pilot on a small number of industries commencing in summer 2016 as we aim to cease MBS for the identified Bands in the appropriate industries and switch to using HMRC turnover data.

## 4 Experimental HMRC turnover data

### 4.1 The data challenge

The HMRC turnover data presently at our disposal has 4 key challenges:

- cleaning of suspicious data
- apportioning the data to the reporting unit from the IDBR enterprise unit
- deriving monthly data from quarterly and annual VAT returns
- forecasting

The first three of these challenges were described in section 5.3 of our [second article](#). As outlined in that article, the cleaning of this large dataset is a significant challenge and can be understood in three parts – a cleaning rule to identify and amend businesses reporting in thousands of pounds rather than pounds; a second cleaning rule to identify and amend businesses reporting with a quarterly pattern (this includes the same values for four quarters signifying an annual return, same values in three quarters and a fourth return as a balancing item, and three zero returns followed by a positive value signifying an annual return); and a suspicious turnover rule to identify and amend returns that are out of the ordinary.

The amendment of values due to the suspicious turnover rule is an area where further work is required as at present some of the values are not cleaned to a credible path. So at this point we have decided to be consistent in only using and sharing data that have NOT been cleaned using these three rules rather than adopting two cleaning practices while omitting the third.

Secondly, the data we are using at this point, and will share in this article, have been apportioned from VAT reporting unit to reporting units where appropriate using the registered employment of the businesses covered by the VAT return. We will also test this method using other variables including registered turnover as these may overcome industries or examples where the employment method is too simplistic.

Thirdly, we should again make clear that the data we are using at this point and sharing in this article have been moved from, for example, a variety of quarterly reporting staggers to a monthly dataset simply by dividing the returns by three. In future we will consider whether we can instead use calendar days, working days or apply a cubic spline to arrive at monthly data.

The final challenge is to use the data at our disposal to forecast HMRC turnover data for selected industries when we begin to test our new method.

## 4.2 Candidate industries for the use of HMRC turnover in summer 2016

For the pilot we have selected some 20 candidate industries for analysis. Most of the selected industries have displayed good coherence when comparing quarter on quarter turnover growth rates between MBS and the existing HMRC turnover data (even with the challenges described above) while some have also been included where Band 1 turnover is relatively small when the Band 1 sample of businesses appears relatively large.

It must be noted that we have excluded the manufacturing industries as the MBS for this area also requires export turnover data from businesses. The HMRC data we have access to do not cover export turnover as a separate entity so progress with manufacturing will be delayed to a later phase and explored further in a future article.

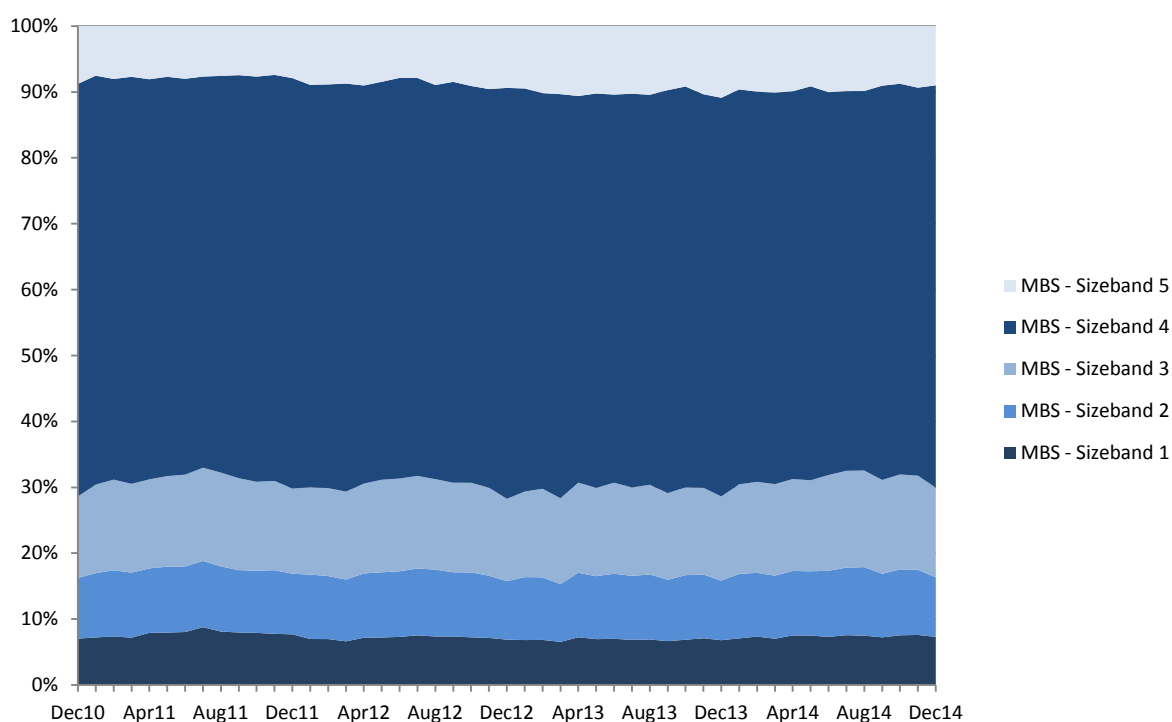
Our aim is to move to using HMRC turnover data for parts of the Index of Services and the Output approach to measuring GDP in summer 2016. We will cease the despatch of MBS forms for the turnover question for small firms in selected industries. In its place we will forecast nominal HMRC turnover data in each appropriate Band before replacing the forecast with an estimate of returned HMRC turnover data at 3 months after the reference month or later, depending on the outcome of our analyses. We will be able to conduct parallel-runs as outlined in section 3.3 and will share the outcomes in future published analyses and articles.

### 4.3 MBS Employment Sizeband analysis

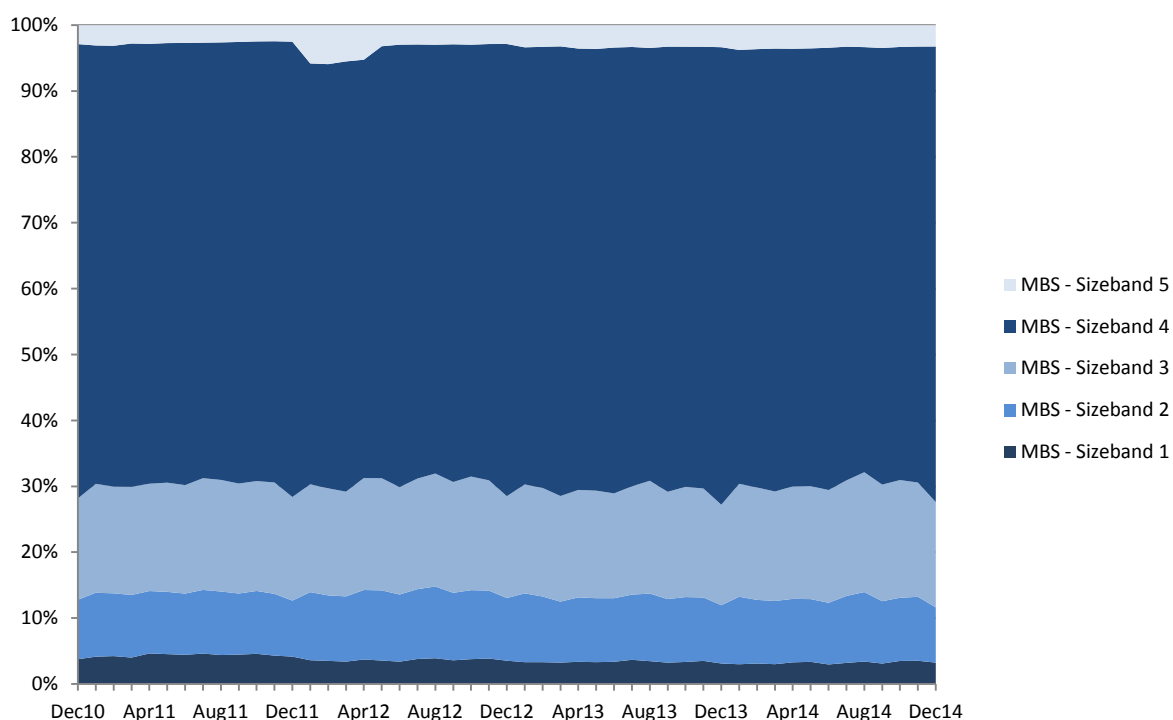
The data within figures 1 to 3 provide a breakdown of the turnover share on a monthly basis for each employment size-band from December 2010 to December 2014 for production and services industries separately and for the overall economy. These data are unpublished at this level. However, the industry level aggregates are published as part of the monthly release [Turnover and Orders in Production and Services Industries \(TOPSI\)](#). All data within figure 1 to 12 are on a non-seasonally adjusted basis.

From figures 1 to 3, two points are evident. Firstly, there is a significant share of the turnover data that is reported in Bands 4 and 5. This is approximately 70% of the overall monthly across the time-series. This is important to emphasise because under the proposed combined MBS/VAT approach this element of turnover would continued to be sampled as part of the MBS. This ensures a more timely return of genuine monthly data which we are able to validate. The other point is the importance and stability of the Bands 1 and 2. These are relatively flat in their profile and have around 15% of the turnover share on a monthly basis across the time-series.

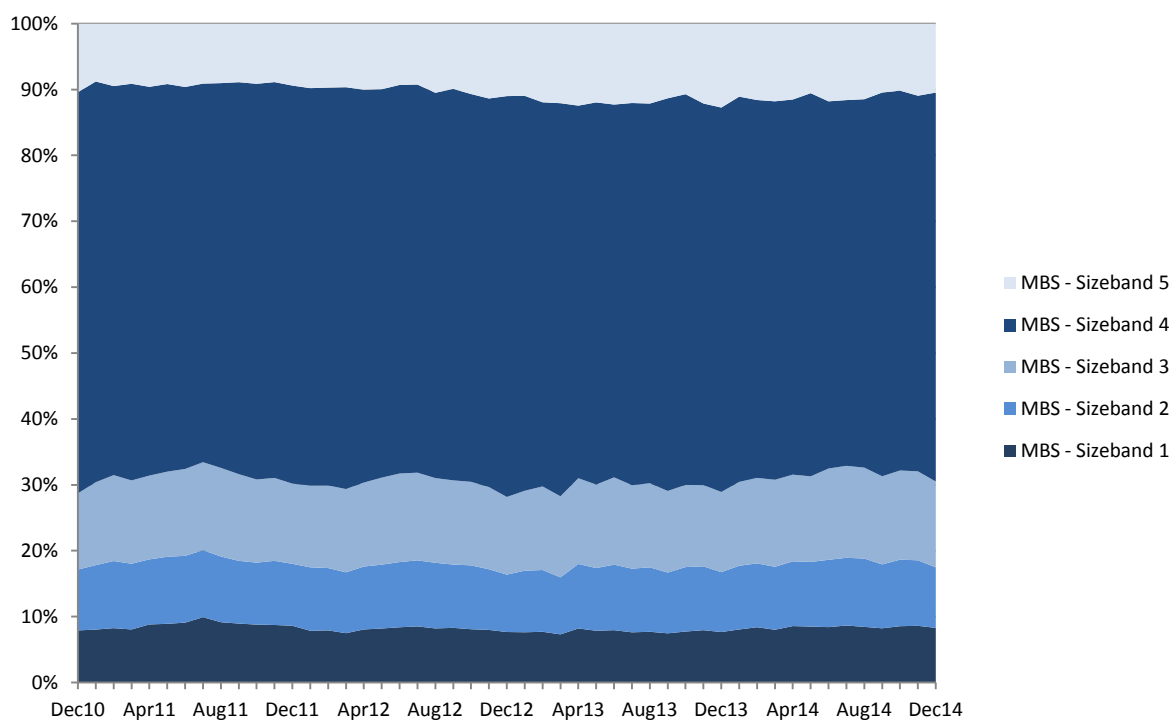
**Figure 1: MBS (All Industries) Turnover – share of turnover by employment size-band**



**Figure 2: MBS (Production) Turnover – share of turnover by employment size-band**



**Figure 3: MBS (Services) Turnover – share of turnover by employment size-band**





#### 4.4 A sample of the candidate industries – comparing HMRC and MBS turnover for SIC2007 industries 55 Accommodation, 58 Publishing activities and 75 Veterinary activities

Using a similar methodology as illustrated within 4.3, the data in figures 4 to 9, replicates the MBS employment size-band analysis for HMRC turnover data. This is for a selection of industries which have illustrated good congruency based on the quarter on quarter turnover growth rates. As highlighted in section 4.2, these are a sample of the candidate industries which are being analysed to see whether they could form part of the pilot for 2016. Please note that MBS Band 5 forced selection businesses feature in their appropriate employment Band for HMRC as these are no longer forced selections.

We should add that all data published in figures 4 to 9 are experimental. The data have been matched from VAT reporting unit to its corresponding IDBR enterprise and then apportioned to its corresponding reporting unit based on reporting unit employment share within the IDBR enterprise. The data have also been assigned to their corresponding months based within a VAT return in order to produce monthly data from the quarterly and annual VAT returns. However, it is important to note it is not a finalised dataset due to the issues highlighted with section 4.1. The data used here have not been subjected to any of the ONS cleaning processes.

The selected candidate industries which are highlighted further below are:

SIC 2007 – Division 55 Accommodation

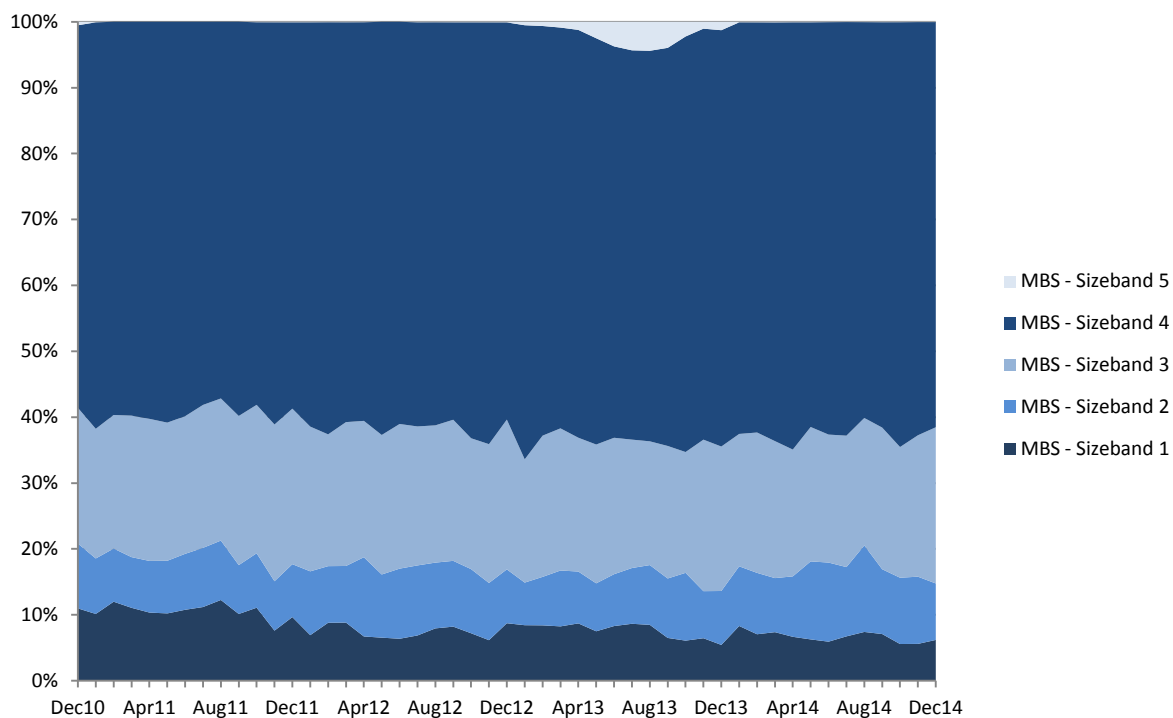
SIC 2007 – Division 58 Publishing activities

SIC 2007 – Division 75 Veterinary activities

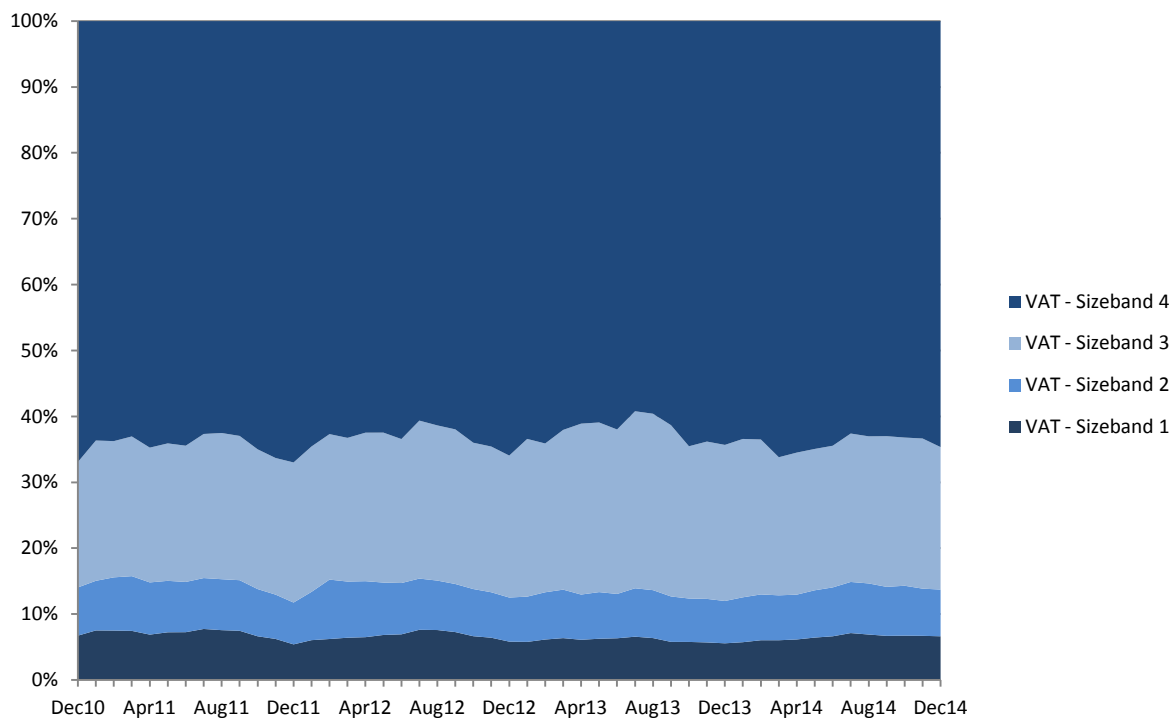
For Accommodation (division 55) and Publishing activities (division 58) the data is similar to the data in figures 1 to 3 as the large share of the MBS turnover is held by the Band 4 business and this is replicated within the HMRC turnover data across the period. Again the Band 1 and 2 businesses make up a relatively small amount of the industry turnover total and this is illustrated in both the MBS and VAT datasets.

Veterinary activities (division 75) is different in character from the data in figures 1 to 3. Here the majority of turnover is reported by small and medium sized businesses and this is replicated through MBS and HMRC turnover. This is therefore a good example of where whilst the smaller and medium businesses dominate the industry in terms of turnover share, both HMRC and MBS turnover act in a similar manner.

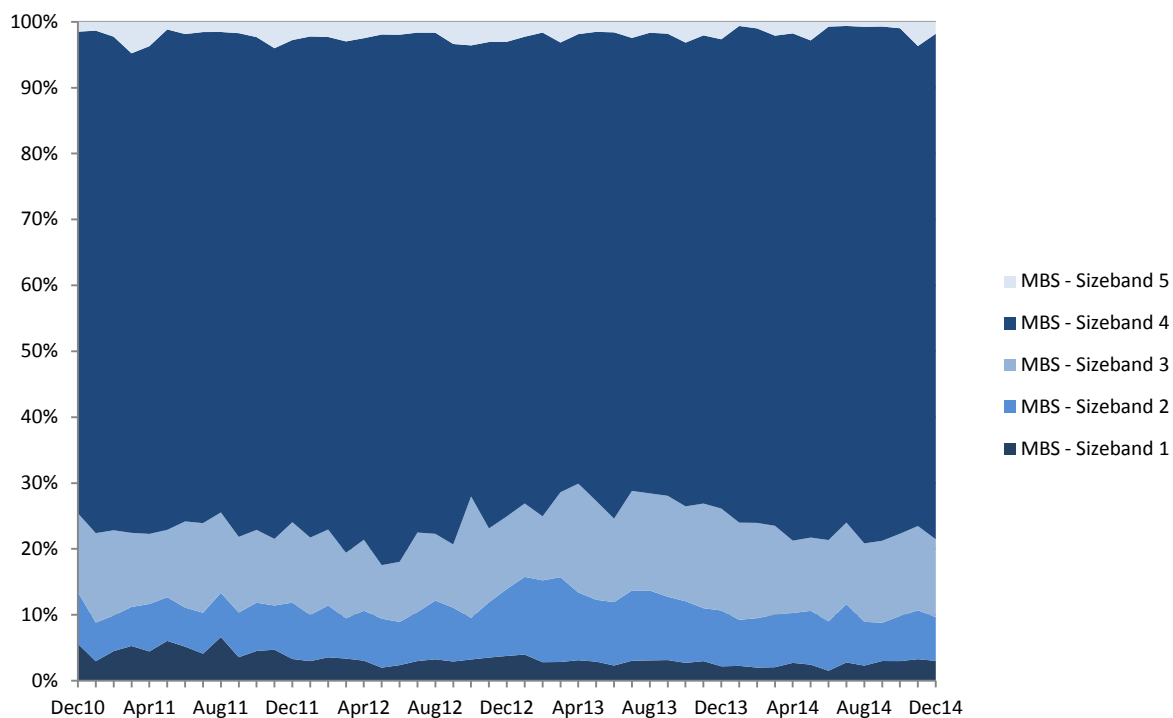
**Figure 4: MBS (SIC 55 – Accommodation) Turnover – share of turnover by employment size-band**



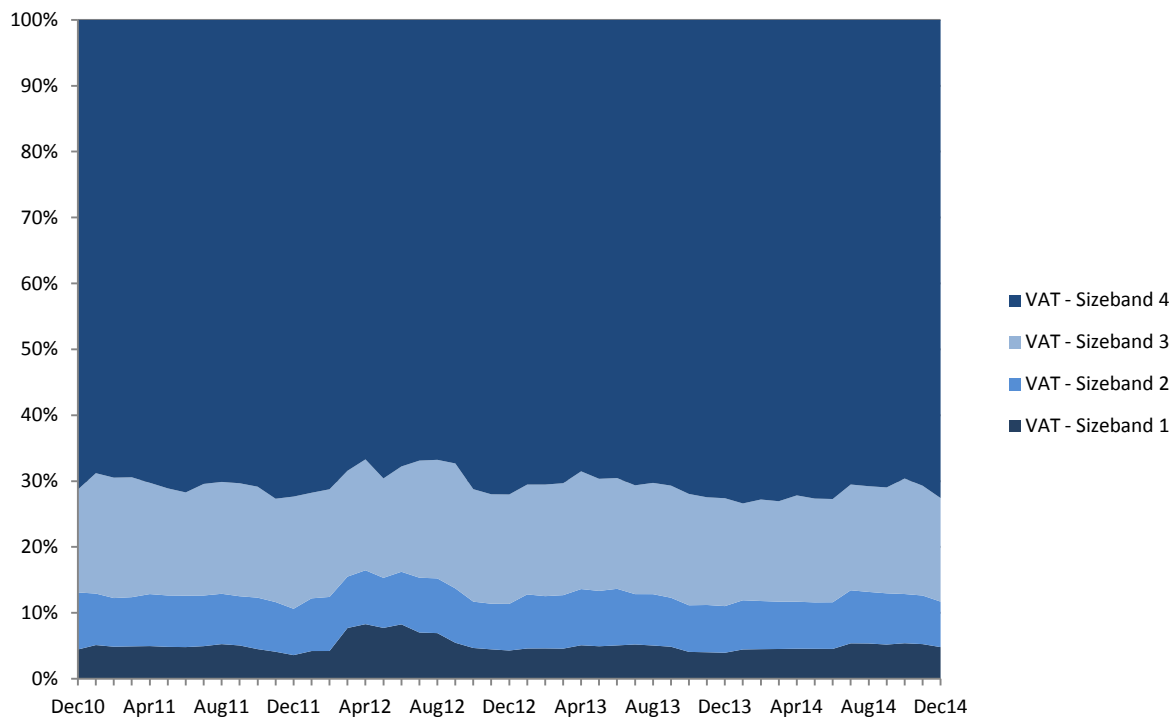
**Figure 5: HMRC VAT (SIC 55 – Accommodation) Turnover – share of turnover by employment size-band**



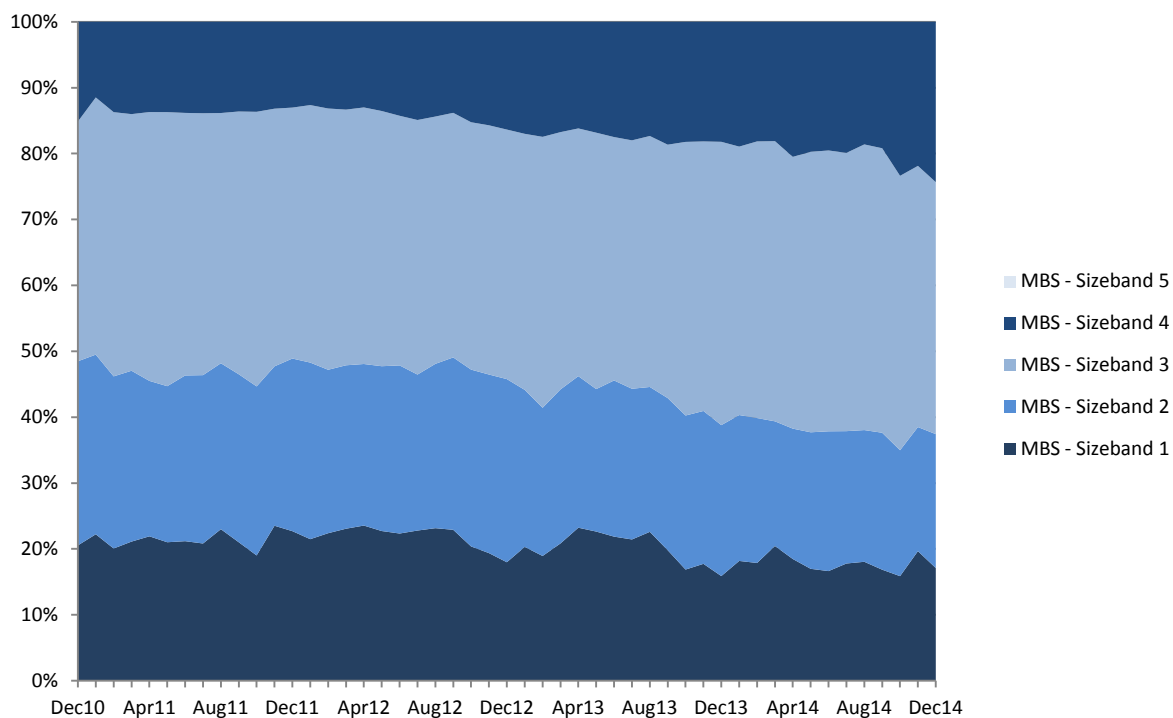
**Figure 6: MBS (SIC 58 – Publishing activities) Turnover – share of turnover by employment size-band**



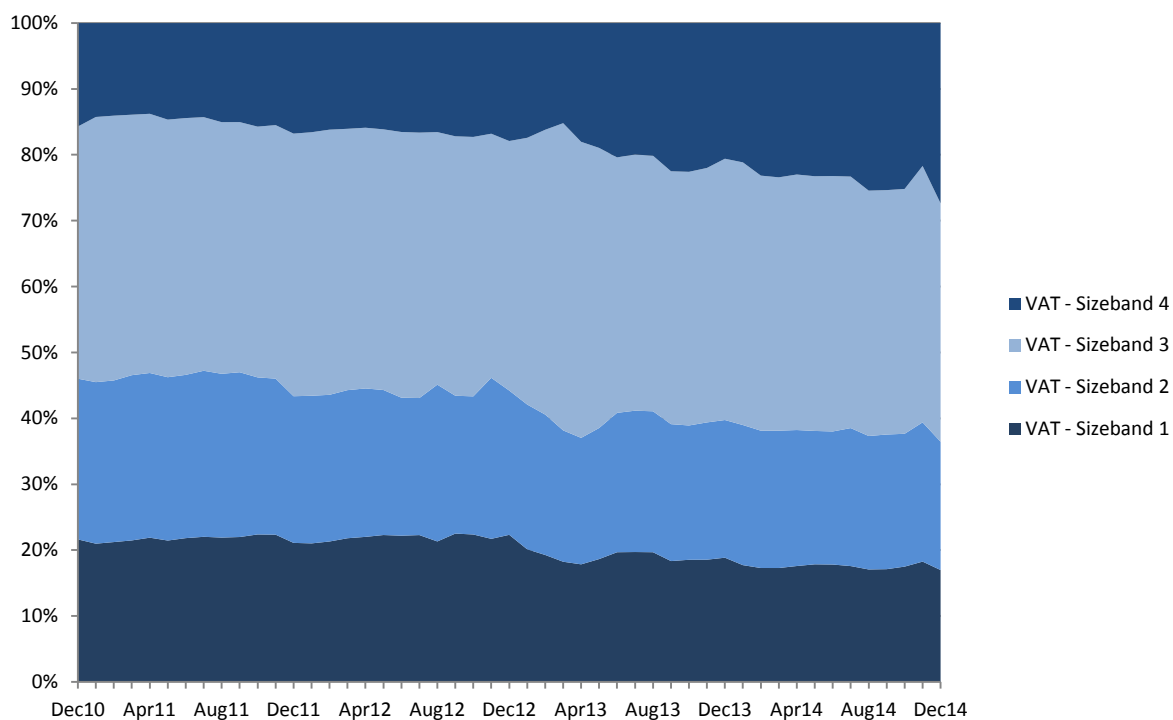
**Figure 7: HMRC VAT (SIC 58 – Publishing activities) Turnover – share of turnover by employment size-band**



**Figure 8: MBS (SIC 75 – Veterinary activities) Turnover – share of turnover by employment size-band**



**Figure 9: HMRC VAT (SIC 75 – Veterinary activities) Turnover – share of turnover by employment size-band**



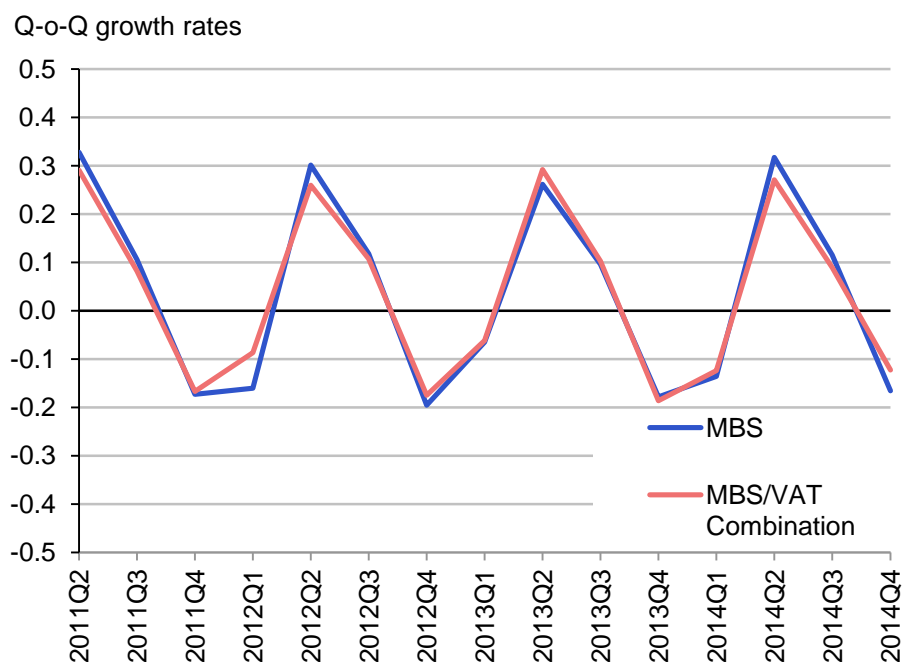
#### 4.5 Quarter-on-Quarter turnover growth rates analysis for the candidate industries

Figures 10 to 12 analyse how the quarter on quarter turnover growth rates would change if a combined MBS and VAT approach were to be used as the method to calculate the turnover series for the industry as explained in section 3.3. This is important to understand as a crucial feature of effective short-term output indicators is to measure growth in the short-term. It should be understood that this is for indicative purposes only and does not indicate that it would be the approach for these industries that we would necessarily adopt.

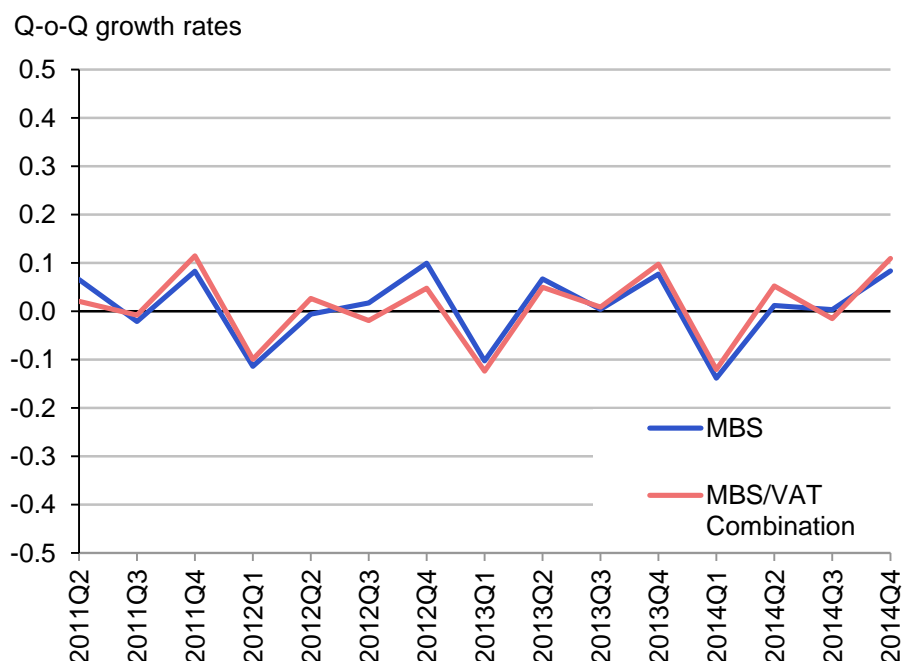
The current quarter on quarter growth as stated by MBS is displayed via the blue line for the time-series Quarter 2 (April to June) 2011 to Quarter 4 (October to December) 2014. The red line has been calculated taking the nominal values from the MBS for the Band 4 and 5 businesses and replacing Bands 1 to 3 with the HMRC turnover nominal series. Again, please note that the HMRC turnover data have been calculated on experimental datasets which have not been cleaned as explained in sections 4.1 and 4.4.

It is evident that these three industries display good congruency in their growth rates over time if this new method were to be adopted. One important feature to note is that the MBS/VAT combined series has been calculated using “mature” HMRC turnover data - HMRC turnover which is over 6 months from the reference period month. How the HMRC turnover data developed in the months immediately after the reference period is an area of work which we will pursue in future through parallel-runs. In this context it should be noted that the continuing regular return of MBS data for the Band 4 and 5 businesses is often a significant proportion of the data (see figures 1 to 9).

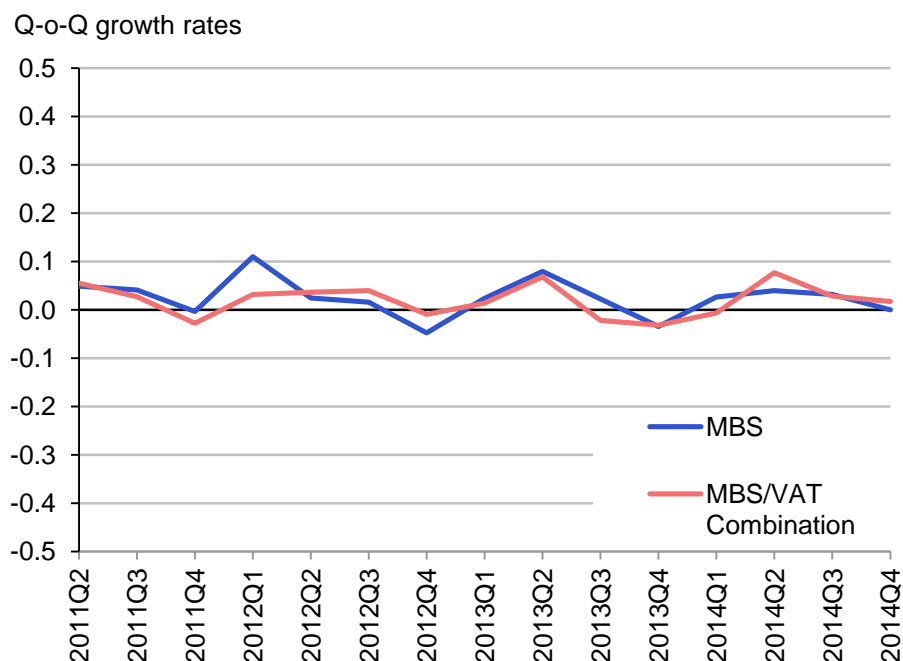
**Figure 10: Quarter-on-Quarter turnover growth rates SIC 55 – Accommodation – MBS in comparison to the proposed MBS/VAT combination method**



**Figure 11: Quarter-on-Quarter turnover growth rates – SIC 58 – Publishing activities – MBS in comparison to the proposed MBS/VAT combination method**



**Figure 12: Quarter-on-Quarter turnover growth rates – SIC 75 – Veterinary activities – MBS in comparison to the proposed MBS/VAT combination method**



#### 4.6 Coverage of businesses – MBS forms in comparison to VAT returns

Tables 3 to 5 illustrate the number of MBS forms despatched (as per October 2015) in comparison to the number of businesses with HMRC turnover data at the points t+3 months and t+6 months from the reference period for the sample of candidate industries (SIC2007 Divisions 55 Accommodation, 58 Publishing activities and 75 Veterinary activities).

Whilst HMRC turnover is less timely in comparison to MBS data it is worthwhile to highlight the significant amount of businesses which report HMRC turnover data at both 3 and 6 months after the reference period end. At t+6 months from the reference period, analysis has shown the dataset to be mature with a high percentage of data returned (see figures 1 and 2 as part of the [first article](#)). This is advantageous because rather than estimating the remainder of a cell universe based on the sample which has responded, we would be able to use actual data as collected via a VAT return. This increased coverage is particularly evident for the smaller employment size-bands (Bands 1 to 3 included). As it is the intention to always retain the Band 4 and 5 businesses the direct comparison for the number of HMRC units has not been provided.

**Table 3 – MBS forms despatched (as per October 2015) for a sample of the candidate industries by Bands**

Band	55 - Accommodation	58 – Publishing activities	75 – Veterinary activities
1	47	24	10
2	83	27	10
3	94	49	10
4/5	537	153	28
<b>Total MBS forms despatched for the industry</b>	<b>761</b>	<b>253</b>	<b>58</b>

**Table 4 – Number of reporting units with HMRC turnover at t+3 months from the reference period for a sample of the candidate industries by Bands**

Band	55 - Accommodation	58 – Publishing activities	75 – Veterinary activities
1	1,666	2,194	449
2	776	625	222
3	856	297	193

**Table 5 – Number of reporting units with HMRC turnover at t+6 months from the reference period for a sample of the candidate industries by Bands**

Band	55 - Accommodation	58 – Publishing activities	75 – Veterinary activities
1	6,529	7,539	1,841
2	2,945	1,763	693
3	3,004	625	549

For example for Division 55 Accommodation, the MBS despatched 47 forms as at October 2015 to businesses within Band 1. If the comparable dataset from HMRC dataset was used this number would increase to 1,666 reporting units at 3 months from the reference period end and would further increase to 6,529 reporting units at 6 months from the reference period end. This significant improvement in coverage over the MBS is not only evident to these industries but across the economy.

The other important point to highlight is that whilst the MBS sample sizes are minimal in size to the number of units eventually captured by the HMRC turnover dataset, the share of turnover for the candidate industries highlighted in the paper by employment size-band is very similar to what is produced in the MBS. For example, for Division 75 – Veterinary activities, for employment sizeband 1 the turnover share of the industry is very similar between HMRC turnover and MBS, even though this achieved with a sample of only 10 reporting units within the Band (see figures 8 and 9). Again this is not evident to only these industries highlighted but in many other industries across the economy.

## 5 Further progress with HMRC data

### 5.1 Regional and expenditure data

In previous articles we have explored the value of a regional analysis of HMRC turnover data and the availability of HMRC expenditure data and both have recently been provisionally delivered for the first time by the project although they have not been tested.

Regional data is available at level 1 of the [Nomenclature of Units for Territorial Statistics \(NUTS\)](#) which is consistent across the EU. This is sourced at reporting unit level but we intend to attempt the analysis at local unit level in the coming months which will improve the regional detail. We are currently focusing our efforts on HMRC turnover for the pilot but will begin to explore regional data in the first half of 2016. When further progress is made with the data challenges described earlier we aim to share the data with users.

The expenditure data has only become available and has not been sense checked. It will be less of a priority in early part of next year but will rise in importance in second half of the year when our work on the turnover pilot will begin to diminish. We will again update users on progress with this data in future articles.



## 5.2 Construction project update

On 11 December 2015 we published a [Construction Statistics Development Programme work plan](#). This included a section on nominal data:

“In early 2016, ONS will work on quality assurance of the nominal data for construction output, balancing between the MBS, the source of nominal data for output in the construction industry and the annual business survey (ABS) but also through reviewing administrative measures such as turnover data provided on VAT returns. The results of this analysis will be published by the end of Quarter 1 (Jan to Mar) 2016.”

It should be noted that the ONS Construction survey asks for “value at work” which is different from the turnover question commonly asked by MBS in other sectors, and of course different from HMRC turnover data. We will continue to inform users of progress in this area.

## 6 Next steps

Over the coming months we aim to overcome the data challenges outlined in the article – improved cleaning of suspicious data, the apportionment of data from VAT units to reporting units perhaps using variables other than employment, and the identification of methods to improve the monthly nature of the data. In addition, we will begin to address a number of issues around the particular characteristics of the HMRC turnover data that will determine how it will be used in future. This will include an analysis of data at business level which will allow us to more closely understand and explain the differences in data reported to MBS and HMRC. A further article to update users on progress with the project is planned for publication in March 2016.

## 7 Contact details

Name: John Allcoat

Email: [stoi.development@ons.gsi.gov.uk](mailto:stoi.development@ons.gsi.gov.uk)

Telephone: 01633 456616

## 8 References

Allcoat, J (2015) "[Feasibility study into the use of HMRC turnover data within Short-term Output Indicators and National Accounts](#)" Office for National Statistics

Her Majesty's Treasury (2015) [Spending Review and Autumn Statement 2015](#), HM Treasury: London

Her Majesty's Treasury and Cabinet Office (2015) [Independent review of UK economic statistics: interim report](#), HM Treasury: London

Stephens, M and Allcoat, J (2015) "[Exploitation of HMRC VAT data](#)" Office for National Statistics

Taylor, C, James, G and Pring P (2011) "[The development of the Monthly Business Survey](#)", Economic and Labour Market Review Volume 5 Number 2, pp 95-103